A MODEL FOR IMAGE PATCH-BASED ALGORITHMS

Karl S. Ni, Truong Q. Nguyen

ECE Dept, UCSD, La Jolla, CA 92093-0407
http://videoprocessing.ucsd.edu/

ABSTRACT

An empirical study of the domain of patch-based learning algorithms for image and video processing is conducted. As patch-based algorithms are commonly used, knowledge of the properties of fixed size image patches would prove particularly useful and interesting. We are concerned with investigating the overall distribution of vectorized patches of general images. A multivariate distribution model is proposed and analyzed using various techniques, which include univariate histograms and modified k-nearest neighbors. The model is verified and an application using the distribution model is introduced and compared.

***Index Terms*—** image patch, distribution, Gaussian Mixture Model, multivariate Laplace, image properties

1. INTRODUCTION

Most problems in image and video processing relate to methods of manipulating, constructing, or correcting new detail from a discrete set of known points. Such problems are usually ill-posed, where the quality of the solution is considered subjectively. Subjective performance focuses on several areas, including but not limited to temporal and spatial continuity, visual comfort with respect to edges and texture, and clarity of content. Rather than using global image properties (of which there is often little correlation between randomly selected images) or processing an entire image at one time, many techniques focus their efforts on achieving these attributes by observing *local* image properties.

Patch-based algorithms operate locally in exactly this manner, and are defined by the processing done in finite dimensional windows consisting of a contiguous subset of pixel locations within a single image. This subset is considered separately from the rest of the image, which in turn contains additional subsets at different locations that are also processed individually.

Because many image/video compression/processing algorithms find the notion of patch-based processing important (e.g. $N \times N$ DCT compression, motion-estimation, etc., are block-based, while postprocessing techniques such as interpolation are windowed), our study concentrates on the properties of the patches themselves, particularly focusing on their *distribution*. Anticipating the usefulness of such an investigation, especially when preprocessing in an input space, several papers [1, 2, 3] have explored this topic. The range of their investigations are diverse, but all of the studies eventually agree on some sort of high-dimensional manifold that cannot be easily explained. [1] assumes two discrete manifolds, termed implicit and explicit, consisting of edges and texture. Likewise, [2] describes image patch space as extremely sparse with data clustered on subspace manifolds, and [3] utilizes synthetic data to offer insight. Our work will explore intuitive explanations behind the manifolds, and why they are broken up as such.

The paper is divided as follows. Sec. 2 proposes a model by considering several works and observations. Next, Sec. 3 introduces a few tools that aid in explaining the model. Sec. 4 provides the necessary parameter estimation steps to use the model. Finally, Sec. 5 runs several experiments involving an applications assuming the proposed model.

2. A UNIFYING MODEL FOR GENERAL PATCH-BASED DISTRIBUTION

Numerical statistics of image patches depend loosely on the size of the evaluation window or the picture shot depth, which can often be associated with the size of the image itself. For example, 7×7 windows taken from the same sequence at different video sizes (QCIF versus CIF versus HD at $720p$ and $1080p$) have different statistics. The situation is similar to appropriately scaling local statistics to correspond to a camera that zooms in on a particular scene. However, the extent of this effect is taken to be minimal, and given a sufficiently large window, [4] suggests approximate scale invariance. That is, if $I(x, y)$ be the observed image, then any given patch of I, $patch(I(x, y))$, approximately reflects the statistics of $patch(I(\sigma x, \sigma y))$.

An important measure in data analysis is the *kurtosis*, or normalized fourth order cumulant, of a distribution. When the normalizing value is 3, then the kurtosis of a random variable X becomes the classical metric of the *Gaussianity* of X.

This work is supported in part by a grant from Qualcomm, Inc., and matching funds from the U.C. Discovery Program

Under Gaussian normalization, the definition is given in (1).

$$kurt(u) = \frac{\sum_i^N (u_i - \overline{u})^4}{(N-1)\sigma^4} - 3, \quad (1)$$

Conventional modeling trends [2] assert that image statistics are seldomly distributed as Gaussians. Instead, any one-dimensional linear mapping of multivariate patch-based distributions, which includes transforms (DCT, DWT, DFT, etc.), display strong *super*-Gaussian, i.e. $kurt(\mathbf{x}) > 0$ (sharply peaked), characteristics. One illustration from the extensive collection of domain modeling includes [5], which uses the wavelet detail domain. The histogram is modeled as

$$X \sim \xi_0 e^{-(|h(\mathbf{x})_i|/\alpha)^\beta}. \quad (2)$$

for any i where $h(\mathbf{x})_i$ is the i^{th} wavelet detail coefficient. (2) results from independently distributed characteristics between detail coefficients, and overall, the proposed model assumes the same in the image patch domain.

The overall form of (2) is eventually important to the proposed distribution model, but in terms of the entire multi-dimensional distribution, a single super-Gaussian peak is only partially descriptive. There are limits on the amount of information a single dimension can convey, which is augmented by most models' almost universal zero-mean properties. Alternatively, patched-based applications, e.g. superresolution [6], often model the overall distribution as a mixture of Gaussians (GMM). Rationalization behind such modeling is seldomly justified in any imaging application, the mindset being more default than deliberate.

However, as will be further clarified in Sec. 3, our belief is that the GMM is a surprisingly accurate model in non-smooth areas, and the proposed model combines a GMM with a single super-Gaussian peak similar to [5]. Thus, we propose a multivariate distribution model of the form:

$$X \sim \pi_0 Q\left(-(\mathbf{x}^T\mathbf{x}/\alpha)^\beta\right) + \sum_{i=1}^{C} \pi_i \mathcal{G}(\mathbf{x}, \mu_i, \Sigma_i), \quad (3)$$

where $Q(\mathbf{z})$ is some kernel (potentially including a modified Bessel function for multivariate Laplace distributions), $\sum_{i=0}^{N} \pi_i = 1$, and $\mathcal{G}(\mathbf{z}, \boldsymbol{\mu}, \Sigma)$ is a multivariate Gaussian PDF.

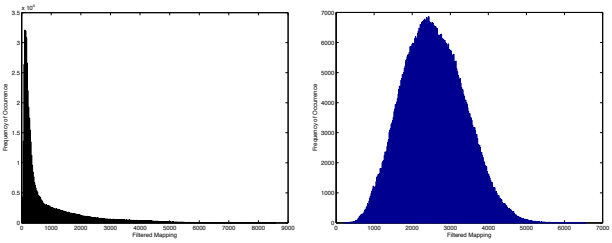
3. EMPIRICAL JUSTIFICATION

There is an extensive quantity of literature on density estimation techniques under various frameworks. The goal of this section is to observe several statistical frameworks that would not only rationalize a parametric method but also verify the functional form of (3).

The simplest density estimation technique is the histogram, and we use the simple mapping as follows for visualization purposes. With $\mathbf{x} = patch\,(I(i,j)) \in \mathbb{R}^{d^2}$, a mapping $u \in \mathbb{R}$, is given in (4).

$$u = \sum_i |\mathbf{x}_k - \overline{\mathbf{x}}| \quad (4)$$

(The mapping is not entirely linear, but for our purposes is sufficient.) [2]'s contention of one-dimensional mappings always having $kurt(u) > 3$ is, in general, true, as seen in Fig. 1(a), where $kurt(u) = 3.203$. Nevertheless, when the image consists of considerable amounts of texture like Fig. 1(d) the histogram will have $kurt(u) \approx 0$ and the Gaussian shape of Fig.1(b).



(a) Histogram of Doorhinge (b) Histogram of Grass

(c) Doorhinge Image (d) Grass Image

Fig. 1. Histograms (a,b) of a 1-D mapping from a 7×7 collection of image patches and the corresponding original images (c,d). Trends in (a) and (b) are typical because (c) contains smooth-objects with edges while (d) is mostly texture.

Explanations behind behavior seen in Fig. 1 are quite intuitive. Small values of u correspond to low "variance" among pixel values in a particular patch $\mathbf{x}$. These areas are the smooth portions or slowly varying gradients of an image, abundant in Fig. 1(c) and images in general. Mathematically, vectors "similar" to $\mathbf{0}$ appear in high concentrations, where skies, surfaces, solid fabric, etc., are most likely present as large contiguous segments of the image. Compared to other types of vectors, an inordinate proportion of such vectors form the basis of the first term in (3). In fact, the presence of edges is almost unnoticeable in Fig. 1(a), with less than a 8% showing. Conversely, Fig. 1(b) exhibits potential for parametrization as $kurt(u) = -0.2388$ using the most nonparametric techniques.

The idea is to infer properties of the multivariate distribution from the statistics of the nonlinear projection given in (4) by assuming some level of correlation between the two.

Given the shape and kurtosis of both histograms, the natural generalization is to fit a super-Gaussian distribution to zero-mean low variance vectors and a GMM to the broad spectrum of texture.

Another traditional distribution modeling technique is k-nearest neighbor (k-NN), where density estimation is mathematically expressed as

$$X \sim \frac{1}{N\sigma} \sum_{i=1}^{N} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{\sigma}\right) \tag{5}$$

where $K(\mathbf{x}, \boldsymbol{\mu})$ represents a kernel that integrates to one. Conceptually, (5) places kernels around each training point for a complete picture of the probability density function (PDF). One proposed approach estimates the PDF a little differently by thinking about the distance from a test point to the nearest training point rather than evaluating the distance based on the surrounding training point locations (though the results are the same).

$$k^*(\mathbf{x}, \Omega) = \begin{array}{ll} \text{argmin}_k & \sum_{i=1}^{N} W_i(\mathbf{x}_t, \Omega, k) K(\mathbf{x}_i, \mathbf{x}) \geq \eta \\ \text{where} & W_i(\mathbf{x}_t, \Omega, k) \in \{0, 1\} \end{array} \tag{6}$$

Using (6), the optimal k^* arises from growing k until η is satisfied by the sum of weights. At some point k-NN becomes ineffective because irrelevant training points begin to be associated with $\mathbf{x}$ in order to meet η. Therefore, we let ζ be a maximum limit on k^*. When $\mathbf{x}$ cannot be represented due to insufficient training, then ζ has been reached in (6) or $k^* > \zeta$.

(a) Original Image (b) $|\Omega| = 2 \times 10^6$

Fig. 2. The lighthouse image is shown in (a). In (b), image patches where $k^* > \zeta$, i.e. insufficient training, are shown in white while well-represented patches are in black.

The conclusion drawn from Fig. 2 is that edges are poorly represented by k-NN, and moreover, edge variety is difficult to accommodate in traditional data collection. A training set is, on the other hand, sufficiently representative of smooth patches and texture in a randomly selected image. Observations from (2) are intuitive; only a very small minority (albeit the most important parts) of the image can be considered edge patches. Furthermore, variability among edges is extremely high, where orientation, contrast, and sharpness all must be comparable for the patch to be considered similar. Thus, the conjecture that low variance patches and texture form the super-Gaussian and Gaussian mixture of (3) resurfaces.

4. USING THE PROPOSED MODEL

Conventional applications of mixture models such as [6] depict each component of the mixture as a class of image content that is independent of other components. Algorithms can provide content-based processing based on the class of the mixture. In practice, the proposed model does not differ much from a GMM, where expectation maximization (EM) experimentation usually yields a number of low-variance, high π_0, and near-zero mean Gaussian.

Yet, by modeling $Q(\mathbf{z})$ differently, higher accuracy may be achieved in the initial term of (3). A common super-Gaussian distribution in practice is the multivariate symmetric Laplace distribution, in which it can be shown that $kurt(u) = 3$ in the univariate case. The behavior of multivariate Laplacian distributions is attractive because it belongs to a family of *sparse distributions* where there is a high peak at the mean [7]. Ignoring the extraneous terms that are unnecessary for our model, the distribution is defined in (7).

$$L(\mathbf{x}) = \frac{2}{\left(2\pi\frac{1}{\alpha}\right)^{\frac{d}{2}}} \left(\frac{\mathbf{x}^T\mathbf{x}}{2}\right)^{\frac{1}{2}-\frac{d}{4}} I_{1-\frac{d}{2}}\left(\sqrt{\frac{2\mathbf{x}^T\mathbf{x}}{\alpha^2}}\right), \tag{7}$$

where I_λ denotes the modified Bessel function of the third kind, and it can be shown that the maximum likelihood estimate of α satisfies (8).

$$\alpha^* > det\left(\frac{1}{N}\sum_{i=1}^{C} \mathbf{x}\mathbf{x}^T\right)^{1/d} \tag{8}$$

Eventually, parameter estimation via EM under the assumption of (3) should give rise to the precise solution, but a simpler alternative would be to first use EM under the assumption of GMM, define a hypercube S around $\mathbf{0}$ where training points are relevant to $Q(\mathbf{z})$, and then approximate parameters of $Q(\mathbf{z})$ using only $\mathbf{x}_i \in S$. In doing so, we replace Gaussian curves in or near S, which may or may not be overfitting the distribution near $\mathbf{0}$, with $L(\mathbf{x})$.

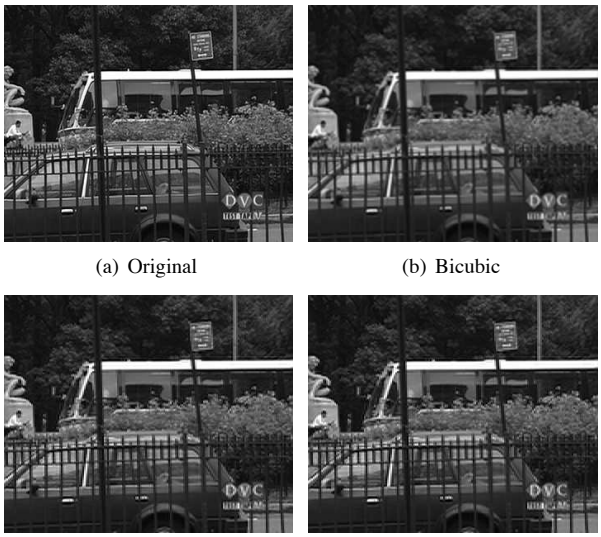
The rationale behind the approximation is that in the initial EM steps, Gaussian curves are placed inside S to accommodate the peaky behavior of the data when in reality, the agglomeration can be attributed to a much simpler phenomenon. Because the Laplace distribution is often described as a multivariate scale mixture of Gaussian models [7], (7) serves as an ideal replacement around $\mathbf{0}$.

5. RESULTS

This section assesses the quality through its performance in various applications. A number of algorithms assume para-

metric models in image processing. One of the simpler applications that illustrates the effectiveness of the proposed model is content-based image interpolation, where [6] is the comparison. Given an image patch $\mathbf{x}$, $\mathbf{x}$ is classified [6] relies on GMM to classify image content,

Following Sec. 4 and using an S with side value 10, we redistribute the π_i values of the eliminated Gaussians using volumes under the curves, coming up with a new $\boldsymbol{\pi}$ as consistent with (3). The results for the GMM interpolation and (3) along with bicubic interpolation are shown in Table 1 and Fig. 3. Table 1 refer to SQCIF, QCIF, and 720p formats.



(a) Original (b) Bicubic

(c) GMM Only (d) Laplacian-GMM

Fig. 3. Visual Results of Classification-Based Approaches

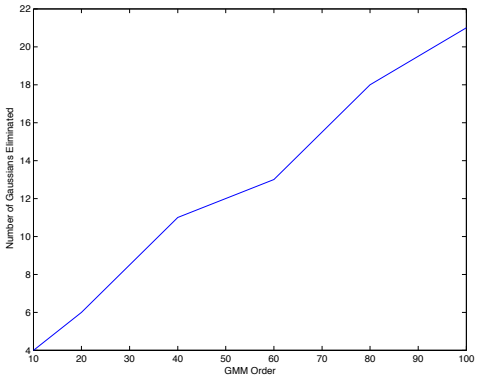


Fig. 4. Properties of Reduction in Model Order

The contention with respect to interpolation is that disparity between the reconstruction efforts of GMM and (3) is negligible. Moreover, as the difference in operation involves vectors near $\mathbf{0}$, meaning there is little or no variance in pixel values within the vectored patch, visual differences are imperceptible. The conclusion in Fig. 3 is that (3) with 15 components achieves the same performance as a GMM of order 20, where a submixture of six Gaussians has been eliminated.

Table 1. Comparisons of Interpolation Model Replacement

	Low-Res	Med-Res	High-Res
GMM-Based	22.6531 dB	22.9722 dB	20.0156 dB
Model in (3)	22.6529 dB	22.9722 dB	20.0156 dB

Consistent clustering around $\mathbf{0}$ occurs quite frequently, and under different training sets, the behavior remained the same. Fig. 4 demonstrates the near linear trend of reduction in model components of (3) as a function of initial GMM components. In replacing the Gaussians, (3) is validated because a mixture of Gaussians with close means is successfully modeled as a Laplace distribution. Generalizing results seen in Fig. 3 to trends in Fig. 4, (3) offers a simpler model than a given GMM with many mixture components around $\mathbf{0}$.

6. CONCLUSIONS

A unifying model of image patches for general use has been proposed. The model is based on observations that images usually consist of a large proportion of low variance image content. Parameters are approximated using the data points in a prespecified hypercube, not the entire training set. The model has been tested in a conventional interpolation framework and shown to be equivalent in quality while simpler in model complexity than the GMM.

7. REFERENCES

[1] K. Shi and S.C. Zhu, "Mapping natural image patches by explicit and implicit manifolds," in *CVPR07*, 2007, pp. 1–7.

[2] A. Lee, K. Pedersen, and D. Mumford, "The complex statistics of high-contrast patches in natural images," 2001, private correspondence.

[3] Ann B. Lee, Kim S. Pedersen, and David Mumford, "The nonlinear statistics of high-contrast patches in natural images," *Int. J. Comput. Vision*, vol. 54, no. 1-3, pp. 83–103, 2003.

[4] Kim Steenstrup Pedersen, *Statistics of Natural Image Geometry*, Ph.D. thesis, University of Copenhagen, Denmark, 2003.

[5] Stephane G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, July 1989.

[6] C. Brian Atkins and C. Bouman, *Classification based methods in optimal image interpolation*, Ph.D. thesis, Purdue University, 1998.

[7] T. Eltoft, T. Kim, and T. Lee, "On the multivariate laplace distribution," *IEEE Signal Processing Letters*, vol. 13, no. 5, May 2006.